

September 11, 1963

Dr. Seymour Benzer
Biophysics Laboratory, Physics Building
Purdue University
Lafayette, Indiana

Dear Seymour:

Some while ago we discussed the computer analysis of overlapping deletion data. At first it seemed like a difficult problem, but Esther had some more data on the Gal system and I recently found a new approach that seems to be generating an efficient program. It hasn't been timed or completely written yet, but I would guess it would come to about $100 n^2$ microseconds of 7090 time to order data on n elements (i.e. about \$10 for an array of 1000) and very probably not 10 X higher than this. The program will operate either ~~dir~~ directly from the recombination matrix, or from genotype lists - the latter especially when a complete $n \times n$ recon test was not feasible and some judgment is entailed in locus assignments. It will produce a test of linear compatibility of the deletions, flag any anomalies, and if these are not too numerous, will produce an ordered map showing any remaining degrees of freedom in the assignments.

I was thinking of going straight to your published diagrams, but it makes more sense to ask you if you have any data sets that you would like to throw directly into the program. Any format on which I could throw an algorithm to generate a recombination matrix would be suitable; probably the best is what is most convenient for you, or what would be most compact to keypunch.

When are you coming out this way?

I have been spending a large part of the past year learning about computers and have just begun to see possible payoffs. I would very much like to talk to you about this, especially some work Emil Zuckerkandl and I are starting on the evolution of protein sequences.

Best wishes,

Joshua Lederberg
Professor of Genetics

JL:edf

S. Benzer